



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/602,262	06/23/2000	Paul S. Cohen	YOR9-2000-0131-US1	9324

7590 03/05/2004  
Anne Vachon Dougherty  
3173 Cedar Road  
Yorktown Heights, NY 10598

EXAMINER

LERNER, MARTIN

ART UNIT	PAPER NUMBER
----------	--------------

2654

DATE MAILED: 03/05/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PRG

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	09/602,262	COHEN ET AL.	
	<b>Examiner</b>	<b>Art Unit</b>	
	Martin Lerner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 22 January 2004.
- 2a) ☐ This action is **FINAL**.                      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1 to 10 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1 to 10 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |   |   |
|---|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)                        | 4) <input type="checkbox"/> Interview Summary (PTO-413)                     |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)    | Paper No(s)/Mail Date. _____  |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| Paper No(s)/Mail Date _____   | 6) <input type="checkbox"/> Other: _____                                    |

## DETAILED ACTION

### *Specification*

The disclosure is objected to because of the following informalities:

On page 11, line 23, "audiovisul" should be —audiovisual—.

Appropriate correction is required.

### *Claim Rejections - 35 USC § 112*

The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

Claims 1 to 10 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claims contains subject matter, which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention.

The limitation of "along with the original video content without altering the original video content" is new matter. Applicants' Specification does not disclose anything expressly about an original video content, or particularly, without altering the original video content. Nor can one having ordinary skill in the art deduce anything implicitly about not altering the original video content from the filed Specification. Applicants

Art Unit: 2654

have not pointed to anywhere in the Specification providing support for the new claim limitations. Apparently, Applicants are improperly attempting to amend their claims in a manner to circumvent the prior art; however, their Specification does not support the claims as now presented. Unaltered original video is not a feature that would be conveyed to one skilled in the art as possessed by the inventors at the time the Application was filed.

***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1, 3 to 5, and 7 to 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.*

Regarding independent claims 1 and 9, *Chen* discloses a sound-synchronized video method and system, comprising:

“processing a video signal to generate a video output comprising at least one time stamped acoustic identification of the content of the audio associated with the video signal along with the original video content without altering the original video content” – codec CD1 separates the digitized video and audio signals into the digital video and speech components; at the video output of codec CD1, a feature extraction module FE1 extracts mouth information visemes containing the mouth shape and mouth

location from the decoded video signal; a memory ME1 stores and time stamps mouth information from the feature extraction module FE1 for phoneme-to-viseme identification (column 2, lines 5 to 47; column 4, lines 36 to 41: Figure 1); according to one embodiment, a viseme is obtained by using a face model to synthesize the mouth area; this is accomplished with a wire frame model (column 4, lines 10 to 25); thus, in this embodiment of *Chen*, the video content is a synthesized wireframe model, so there is no alteration of the original video content;

“processing an audio signal to generate an audio output comprising at least one [time stamped] acoustic identification of the content of said audio signal” – codec CD1 separates the digitized video and audio signals into the digital video and speech components; a phoneme recognition module PR1 divides the incoming speech components into recognizable phonemes; lookup table LT1 maps phonemes into visemes (column 2, lines 5 to 22; column 4, lines 26 to 35: Figure 1);

“synchronizing the video signal to the audio signal by adjusting at least one of the signals to align at least one acoustic identification from the video signal with a corresponding acoustic identification from the audio signal” – video and audio signals that had become unsynchronized are displayed by synchronizing the video frame to produce sound synchronized video (column 4, lines 33 to 63: Figure 2).

Concerning independent claims 1 and 9, *Chen* discloses the video signal is time stamped, but omits time stamping the audio signal. Only one of the audio and video signals is expressly time stamped in *Chen* because visemes are employed as a reference to synchronize the signals. However, it is common in the prior art to assign

Art Unit: 2654

time stamps to both audio and video data streams for purposes of synchronization to an absolute time reference. *Braida et al.* teaches a related method and system for synchronizing video images to speech elements where time stamps are applied to both audio and video streams. Phone recognition program 44 assigns start and stop times to digital speech samples 32 (column 6, lines 53 to 58), and digital video images also have time stamps which are referenced to the same time (column 12, lines 13 to 29). It would have been obvious to one of ordinary skill in the art to additionally apply time stamps to the audio signals as taught by *Braida et al.* in the synchronization method and system of *Chen* for the purpose of providing an absolute time reference for synchronization.

Regarding claim 3, *Chen* discloses phoneme recognition module PR1 produces visemes ("the audio identification") from the audio signal and feature extraction module FE1 extracts corresponding mouth information visemes from lookup table LT1; the output video is applied to display DI1 together with the audio signal and produces lip synchronization (column 2, lines 11 to 38: Figure 1).

Regarding claims 4 and 10, *Chen* discloses a method and system for processing a video image, comprising:

"extracting at least one image from the video signal" – codec CD1 separates the digitized video and audio signals into the digital video and speech components (column 2, lines 6 to 11);

“detecting at least one feature in said at least one image” – a feature extraction module FE1 extracts mouth information visemes contain the mouth shape and mouth location from the decoded video signal (column 2, lines 21 to 39: Figure 1);

“analyzing the parameters of said feature” – mouth deformation module MD1 receives inputs from the video signal and information from the feature extraction module FE1, and visemes from lookup table LT1 (column 2, lines 21 to 39: Figure 1);

“correlating at least one acoustic identification to the parameters of said feature” – a viseme is selected from lookup table LT1 that matches features extracted by feature extraction module FE1 (column 2, lines 21 to 39: Figure 1).

Regarding claims 5 and 7, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes (“articulatory type”) rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2).

Regarding claim 8, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes (“articulatory type”) rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2); *Chen* discloses feature extraction module FE1 extracts mouth information visemes containing mouth shape (“a facial feature”) (column 2, lines 18 to 31).

Art Unit: 2654

Claims 2 and 6 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.* as applied to claim 1 above, and further in view of *Basu et al.* ('885).

Concerning claim 2, *Braida et al.* discloses a Viterbi search for purposes of phone recognition (column 6, lines 59 to 61; column 7, lines 51 to 53), but omits utilizing a Viterbi search for purposes of synchronization. However, it is well known that a Viterbi algorithm is utilized for both recognition and time warping alignment. *Basu et al.* ('885) teaches a method of aligning phonemes and visemes with a Viterbi algorithm. (Column 1, Lines 53 to 67) It would have been obvious to one having ordinary skill in the art to utilize a Viterbi algorithm as suggested by *Basu et al.* ('885) in the synchronization method and system of *Chen* for the purpose of aligning phonemes and visemes more accurately.

Regarding claim 6, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes ("articulatory type") rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2).

### ***Response to Arguments***

Applicants' arguments filed 22 January 2004 have been fully considered but they are not persuasive.



Applicants argue *Chen* does not disclose time stamping of “the original video content without altering the original video content”. Applicants maintain *Chen* is not synchronizing the streamed video signal to the live audio signal, but is replacing or overlaying the live video signal with stored visemes to match the audio. Thus, Applicants say *Chen* alters the original video content, unlike the claimed device, which presents the original video content synchronously with the audio. Also, Applicants argue that *Chen* does not disclose synchronizing the video signal to the audio signal by alignment because *Chen* only covers up non-synchronous live video to appear synchronous. This position is traversed.

Firstly, the amended claim limitations of “an original video content” and, particularly, “without altering the original video content” present new matter. Applicants’ Specification as initially filed does not disclose anything expressly about not altering the original video content. Nor can one having ordinary skill in the art deduce anything implicitly about not altering the original video content from the filed Specification. Applicants have not pointed to anywhere in the Specification providing support for the new claim limitations. Apparently, Applicants are improperly attempting to amend their claims in a manner to circumvent the prior art. Applicants say in their Remarks, Page 8, “Applicants have amended the language of the independent claims to highlight the distinction over the *Chen* approach.” However, their Specification does not support the claims as now presented. Unaltered original video is not a feature that would be conveyed to one skilled in the art as possessed by the inventors at the time the Application was filed.

Secondly, there is at least one embodiment where the original video content is unaltered in *Chen*. At Column 4, Lines 10 to 25, *Chen* discloses that, according to one embodiment of the invention, a viseme is obtained by using a face model to synthesize the mouth area. According to an embodiment, this is accomplished with a wire frame model. The entire face is a synthesized video image of a wireframe model in one embodiment of *Chen*. In this embodiment, *Chen* does not utilize the elements of the speaker's extracted video image to overlay the mouth area. Instead, the entire video image is synthesized and original. Thus, *Chen* does not alter the original synthetic video image in this embodiment.

Thirdly, *Chen* clearly discloses synchronizing the video and audio signals. At Column 4, Lines 55 to 63, *Chen* expressly states, "The invention synchronizes video and audio signals that had originally been acquired as synchronized signals but had become unsynchronized by processing." *Chen* then says delay of the video signal relative to the audio signal can occur in various places during processing. Given the express disclosure in *Chen* of synchronizing video and audio, Applicants' contention, that *Chen* merely "covers up" a non-synchronous live video signal, is not persuasive.

Fourthly, the concept of what constitutes "original" and "unaltered" video is ambiguous and not well defined. If the video is time-adjusted, as disclosed by Applicants' Summary of the Invention, Page 3, Lines 8 to 12, then the video is not unaltered in their invention. Correspondingly, the fact that *Chen*, in some embodiments, may overlay a mouth area on an original video image need not imply that *Chen* is "altering" the original video. The original video in *Chen*, with the possible exception of

Art Unit: 2654

mouth area, is unaltered, but the mouth area in *Chen* is not really video because it is synthetic. The problem of interpretation of what constitutes an "unaltered original video content" arises because these terms are not defined in Applicants' Specification.

Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

Therefore, the rejections of claims 1, 3 to 5, and 7 to 10 under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.*, and of claims 2 and 6 under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.* as applied to claim 1 above, and further in view of *Basu et al.* ('885), are proper.

### ***Conclusion***

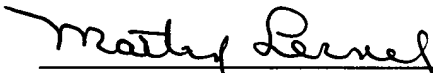
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (703) 308-9064. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (703) 305-9645. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Art Unit: 2654

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ML  
3/3/04

  
\_\_\_\_\_  
Martin Lerner  
Examiner  
Art Unit 2654